Pell Grants as Performance-Based Scholarships? An Examination of Satisfactory Academic Progress Requirements in the Nation's Largest Need-Based Aid Program

Lauren Schudde
The University of Texas at Austin

Judith Scott-Clayton
Columbia University

Pre-print of *Research in Higher Education* manuscript:
The final publication is available at Springer via http://dx.doi.org/10.1007/s11162-016-9413-3

Abstract

The Federal Pell Grant Program is the nation's largest need-based grant program. While students' initial eligibility for the Pell is based on financial need, renewal is contingent on meeting minimum academic standards similar to those in models of performance-based scholarships, including a grade point average (GPA) requirement and ratio of credits completed compared to those attempted. In this study, we describe federal satisfactory academic progress (SAP) requirements and illustrate the policy's implementation in a statewide community college system. Using state administrative data, we demonstrate that a substantial portion of Pell recipients are at risk for Pell ineligibility due to their failure to meet SAP GPA or credit completion requirements. We then leverage the GPA component of the policy to explore the impacts of failure to meet standards on early college persistence and achievement, earning a credential, and transferring to a four-year college using two methodological approaches: regression discontinuity (RD) and difference-in-differences (DD). Our results across the two approaches are mixed, with the RD providing null estimates and the DD indicating some statistically significant impacts, including a negative effect on early college persistence. We conclude by discussing the implications for future research.

Keywords: Higher education policy, need-based aid, persistence, college completion, financial aid, satisfactory academic progress

Need-based financial aid is a widely used tool to promote college attainment, with the federal government delivering $185.1 billion in aid to undergraduates in 2013 (Baum & Payea, 2013). The Federal Pell Grant Program is the single largest source of need-based financial aid for college students in the United States, providing aid to over a third of undergraduates (Baum & Payea, 2013). While eligibility is initially based purely on financial need, recipients must meet satisfactory academic progress (SAP) requirements to maintain eligibility for Pell Grants beyond the first year. Institutions have flexibility regarding how they define SAP but commonly require students to maintain a cumulative grade point average (GPA) of 2.0 or higher and to complete at least two thirds of the course credits they attempt. Those who fail to meet the SAP requirements at the end of the institution's evaluation period may continue to receive aid for one additional term, but if they are still failing SAP at the end of that term, they may lose eligibility for Pell.

The academic requirements for Pell Grant renewal have received little attention from researchers and policymakers, despite their potentially broad consequences. Forty-five percent of Pell Grant recipients entering postsecondary institutions do not enroll for a second year of college (authors' calculations, Beginning Postsecondary Students: 2004/2009). While performance-based scholarships have risen in popularity, the performance-based aspects of the nation's largest need-based grant program have gone virtually unstudied. How many students fail to meet SAP standards? Do the bureaucratic hurdles posed by the standards contribute to the observed low rates of persistence among Pell recipients? The prevalence of SAP failure and its consequences in terms of persistence, achievement, and degree attainment are unknown.

Since SAP policy was first enacted in the late 1970s, enrollment in college has soared. Educational expansion and increasing diversity among college students (particularly in academic preparedness), paired with rising college costs, mean that more students than ever face the

potential repercussions of SAP policy. In light of these trends, there is a clear need for more information about SAP policy and its effects on student outcomes.

National estimates highlight the need to look closer at SAP policies and their impact on student outcomes. As demonstrated in Figure 1, approximately one in five (21 percent) first-year Pell recipients nationally was at risk of losing their Pell due to the SAP GPA criterion alone in 2012. The rate has risen over time since 2004, particularly at public two-year colleges, where nearly a quarter of first-year Pell recipients failed to attain a 2.0 at the end of their first year.

[Insert Figure 1 about here]

In this paper, we estimate the prevalence of SAP failure in a large state community college system. To explore the implications of failing to meet SAP requirements on student outcomes, we use state administrative data to evaluate the effects on early college achievement, persistence, institutional transfer, and degree attainment. We use both regression-discontinuity (RD) and difference-in-differences (DD) approaches to estimate the consequences of failing to meet the SAP GPA requirement. The RD determines the effect of the policy on Pell recipients who just miss the cutoff by comparing them to Pell recipients who just make the cutoff, while the DD determines the effect of falling below the cutoff for Pell recipients (including those further below the cutoff, rather than just at the margin) by comparing Pell recipients to students who are not subject to SAP standards (those without federal financial aid).

To our knowledge, this is the first empirical examination of SAP policy since the 1980s and the first to use multi-institutional data. This paper makes several contributions. First, we clearly outline the federal guidelines for SAP evaluation by postsecondary institutions receiving federal financial aid and describe how evaluation is implemented in a state community college system. Second, we theorize how SAP policy may impact student outcomes, drawing from prior

research on performance-based aid and academic probation. Third, we provide empirical evidence regarding the prevalence and consequences of SAP failure in a statewide community college system.

Approximately 25 percent of first-year Pell recipients fail to attain a 2.0 GPA in our statewide community college sample, in line with our estimates from national data. After accounting for the credit completion requirement, the first-year SAP failure rate approaches 40 percent. Findings on the consequences of SAP failure are somewhat mixed. The RD analysis does not yield statistically significant results, while the DD estimates suggest that failing SAP discourages persistence after the first year of college. We conclude with a discussion of the assumptions of each technique and the implications of results.

Policy Background

Origins of Satisfactory Academic Progress and Federal/Institutional Roles

In 1976, the term "academic progress" was introduced into federal financial aid legislation as an amendment to the Higher Education Act of 1965. The amendment stipulated that, to receive financial assistance under Title IV, students must demonstrate "satisfactory progress" toward a degree (Bennett & Grothe, 1982). The legislation did not explicitly define satisfactory progress, but the Interim Final Regulations in 1978 stated that an institution should evaluate progress by establishing a "normal time frame" for completing the course of study and by measuring grades or work "against a norm" (Bennett & Grothe, 1982, p. 1). Since then, the policy has evolved to be more explicit in its terms while still allowing institutions to determine, within certain confines, how best to evaluate academic progress.

Today, students receiving federal financial assistance must be evaluated at the end of each academic year by their institution. An institution's definition of SAP must align with its

standards for graduation and be "at least as strict" as the academic policy applied to students not receiving assistance from Title IV programs (Satisfactory Academic Progress, 2013, p. 546; Student Eligibility, 2013, p. 571). The language of the federal policy offers some flexibility, but institutions must set a "qualitative standard," in which students meet a cumulative GPA of 2.0 or its academic equivalent by the end of their second academic year; a "quantitative standard," in which there is a "minimum percentage of work" successfully completed; and a "maximum timeframe" standard, in which students finish a program within 150 percent of its published length (Satisfactory Academic Progress, 2013, p. 546).

Students who do not meet an institution's SAP standards become ineligible to receive federal financial aid. However, the federal legislation includes an appeals clause, allowing institutions to waive the provisions of the SAP policy if a student faces "undue hardship," such as the death of a relative, personal injury or illness, or "any special circumstances as determined by the institution" (Student Eligibility, 2013, p. 571).

Rules in Effect for Students in State Administrative Sample

In this study, we explore the application of SAP guidelines within a subset of colleges within a state community college system that evaluates student progress at the end of each term. To meet SAP requirements at the end of each term, students needed to earn a cumulative GPA of 2.0 or higher, complete 67 percent of all credit hours attempted, and finish their program within the maximum timeframe. SAP guidelines were included in students' financial aid letters, published in each college's course catalogs and financial aid office webpages, and, at some institutions, available on "quick facts" handouts provided by the financial aid office.

At the end of each term, colleges notified students of their SAP status. Students who met the minimum SAP standards were deemed satisfactory, while students who had not earned the

required GPA or credit completion rate were placed on warning for the subsequent semester. Students on warning were able to receive aid but risked financial aid termination if they continued to fail to meet SAP standards. Upon failing to meet standards during the warning term, students could appeal to continue receiving federal and state financial aid. Financial aid was terminated for students who did not appeal or made unsuccessful appeal attempts. Students with successful appeals entered a probation period of one term. Students on probation-by-appeal who failed to meet SAP standards subsequently lost aid eligibility.

Theoretical Perspective and Previous Literature

Theoretical Perspective

Informed primarily by the economic literature, we posit that SAP standards affect student behavior through three possible mechanisms. First, SAP policy may provide students with a clear *signal* about their own performance relative to institutional expectations for success. Manski (1989) and Altonji (1993) described college as a process of experimentation and learning about one's abilities. Without a clear signal such as that provided by SAP, some students may persist with a low level of performance, not realizing it will prevent them from graduating. Receiving this signal may induce some students to drop out. Second, as shown by Bénabou and Tirole (2002) in a simple principal-agent model (with parents or policymakers as principals and students as agents), setting minimum performance standards creates *incentives* for individuals near the performance threshold to increase effort, while it creates incentives for some students far below the standard to drop out. The effects of such of policy will, therefore, vary across individuals, and the overall benefit will depend on the magnitude of these competing effects. Finally, even if individuals do not respond to SAP's academic incentives, the loss of financial aid for those who do not improve performance increases the *price* students pay for continued

enrollment, inducing additional dropout among students who face credit constraints and/or conclude that the costs of continued enrollment no longer justify the benefits.

Previous Literature

To our knowledge, there is no prior empirical work evaluating how SAP requirements impact students receiving need-based aid. Extant research on SAP policies is sparse, dated, and limited to descriptive studies of policy implementation within individual institutions (e.g., Bennett & Grothe, 1982; McNair & Taylor, 1988). Nonetheless, we gain helpful insights from related work examining how students respond to academic probation policies, performance-based financial aid, and financial aid receipt in general.

Academic Probation. The prior study arguably most similar to this one is an examination by Lindo, Sanders, & Oreopoulos (2010) of how students respond to being placed on academic probation. Even students not receiving financial aid still typically face the threat of academic probation, particularly at four-year colleges, if they fall below a minimum GPA threshold. Those who fail to improve by the end of the probationary period may no longer be permitted to register. Lindo, Sanders, & Oreopoulos (2010) draw upon the Bénabou and Tirole (2002) model to show how such policies can generate either positive or negative effort incentives, depending upon how far a student is from meeting the standard. The authors use a regression-discontinuity (RD) design on administrative data from a public four-year institution in Canada to compare students just above and below the minimum GPA threshold at the end of their first year. They find that students just below the threshold who are placed on academic probation are less likely to return the next year, but that those who do return improve their grades. A more recent RD study of academic probation policies at a public four-year institution in the United States found similar patterns (Casey, Cline, Ost, & Qureshi, 2015). The findings

also suggested that short-term improvements in GPA appear attributable to strategic course-taking, rather than increased effort—probationary students attempted fewer credits and enrolled in easier courses, (Casey, Cline, Ost, & Qureshi, 2015).

SAP standards have many similarities to academic probation standards. Students on financial aid, however, may be even more sensitive to these types of performance standards given the financial consequences. An additional limitation of prior work on academic probation is that while theory predicts such standards should have heterogeneous effects depending upon how far students are from the threshold, the RD methodologies used in prior work only examine effects at the threshold. Our context will allow us to employ both an RD as well as a difference-in-difference design (comparing students on financial aid, who are subject to SAP, with students not on financial aid), with the latter allowing us to say something about the effects of the policy for students further below the threshold.

Performance-based aid. A number of prior studies have explored the effects of performance-based financial aid. One strand of this literature focuses on large state merit-aid programs that typically require students not only to meet performance criteria in order to receive the scholarship, but to meet academic performance criteria each year to continue receiving aid (Cornwell, Lee, & Mustard, 2005; Cornwell, Mustard, & Sridhar, 2006; Dynarski, 2008). Research on West Virginia's PROMISE scholarship, a state merit-based scholarship program for qualifying high school graduates who maintain a minimum GPA and course load during college, indicates substantial impacts at the requirement thresholds, supporting the notion that academic incentives promote academic effort (Scott-Clayton, 2011). Still, many recipients of merit-based scholarships lose eligibility due to ongoing performance standards and those students who just barely lose aid have worse college outcomes than those who just barely keep it (Carruthers &

Ozek, 2013; Henry, Rubenstein, & Bugler, 2004). Furthermore, while programs like WV's PROMISE illustrate the effectiveness of incentives for a population with demonstrated prior academic success, it is unclear whether such standards are similarly effective when applied to a less-high-achieving population, such as need-based aid recipients.

MDRC, however, has run a number of experiments testing the effectiveness of performance-based scholarships for students receiving Pell Grants. The awards are not conditioned on prior academic achievement and, instead, are supplemental grants (i.e., provided in addition to existing aid) earned over the course of a term by meeting academic benchmarks (Welbeck, Ware, Cerna, & Valenzuela, 2014). Advocates expect to see improved achievement, progress through requirements, and degree attainment as a function of tying aid to performance benchmarks (Mayer, Patel, Rudd & Ratledge, 2015).

The Opening Doors Demonstration in Louisiana was MDRC's first experiment with performance-based awards. Results indicated that the program increased persistence in the second semester and credit accumulation in the first year (Barrow et al., 2014; Richburg-Hayes et al., 2009; study follow-up was curtailed due to Hurricane Katrina). Its positive impacts inspired the launch of the national Performance-Based Scholarship (PBS) Demonstration in 2008, which tested variations of PBS in six different sites across the country. Results from these additional sites indicated improvement in credit accumulation after the first year and little effect on persistence, but modest effects on degree completion (3 percentage points) five years after entry (Mayer et al., 2015).

Taken together, these findings indicate that performance incentives may be effective for students across a range of prior achievement in various contexts. Still, SAP may affect students differently because it is a negative incentive that threatens students' foundational aid, rather than

a positive incentive to receive extra assistance. Separate from the literature on academic standards and performance-based aid, it is helpful to consider SAP policy in light of what we know about the effectiveness of financial aid generally, because students who fail to meet SAP ultimately face the consequence of losing their aid eligibility completely.

Need-Based Financial Aid. Need-based financial aid aims to offset the challenges faced by low-income students by alleviating financial constraints. Financial aid through grants has been shown to improve enrollment, persistence, credit accumulation, and degree attainment (e.g., Bettinger, 2004; Dynarski, 2003; Goldrick-Rab, Kelchen, Harris, & Benson, forthcoming). Evidence is somewhat more mixed in the two-year setting (see Angrist, Autor, Hudson, & Pallais, 2014; Barrow, Richburg-Hayes, Rouse & Brock, 2014; Castleman & Long, 2013; Cohodes & Goodman, 2014; Kane, 1995). However, students at community colleges appear to be particularly sensitive to tuition and discounts (Attewell, Heil, & Reisel, 2011; Denning, 2014; Martorell, McCall, & McFarlin, 2014).

Much of our knowledge about need-based grants focuses solely on the impact of providing money, as though it were offered with no strings attached. But there are bureaucratic hurdles for retaining financial aid, including the academic requirements attached to many awards (Goldrick-Rab, 2013; Goldrick-Rab et al., forthcoming). Given that financial aid aims to offset financial constraints for college students, federal financial aid policy should, at the least, be designed to do no harm to students (Scott-Clayton, 2013). The complexity of navigating financial aid, even after initially receiving the award, may undermine this simple goal. If students are not fully aware of SAP policy, then its effectiveness as an incentive may be undermined. In a context of poor information and communication, the primary effect of SAP requirements may be punitive—simply removing students from aid eligibility—rather than formative.

Research Questions

The theoretical perspectives and prior empirical literature discussed above drive our hypotheses that tying academic standards to need-based aid may both incentivize effort by encouraging students within range of the threshold to meet the standards *and* discourage persistence by weeding out some students who are either unable to meet expectations by the end of the evaluation period or who decide it is no longer worthwhile to try.

To determine the magnitude of SAP failure and evaluate the extent to which SAP standards impact student achievement (which tests the incentivizing hypothesis), persistence (which tests the discouragement hypothesis), as well as transfer and attainment outcomes (related to both processes), we ask a series of research questions:

1. What is the prevalence of failure to meet academic standards in a statewide community college system and how is it changing over time? How many Pell recipients fail to meet the GPA and credit ratio requirements in a given term, and how are these patterns related to student persistence over time?
2. For Pell recipients in a narrow range around the GPA cutoff for SAP, how does failing to meet the GPA standard impact student persistence and student achievement, transfer, and degree attainment?
3. Does failing the GPA standard affect the outcomes of Pell recipients differentially compared to students with the same GPAs who are not subject to the standards (i.e., non-aid recipients)?

Data

To answer our research questions, we use administrative data from a state community college system (SCCS) that includes 58 separate public two-year institutions. We restrict the

sample to the 49 colleges with equivalent SAP policies, obtained from course catalogs published during the sample timeframe. Our available data includes student-level data from first-time college entrants in all SCCS colleges from 2001-2010. We focus on cohorts with at least three years of follow-up data to track enrollment and transfer behavior, as well as degree attainment, which limits us to the 2007 cohort as our final cohort. We were unable to include students from the 2001 cohort, as it was missing financial aid data. The final analytic dataset includes a pooled sample of 147,380 first-time students, 42,835 of whom received Pell Grants, from six cohorts of fall-term enrollees (2002–2007). In our analyses, we pool across the cohorts to maintain our robust sample size, although we control for potential cohort effects, as described below.

The state administrative data track federal and state aid and transcript measures, including courses, credits attempted and completed, and grades for each term. The data are derived from three sources: SCCS data with student information on demographics, transcripts, and financial aid; National Student Clearinghouse graduation data; and Employment Security Commission earnings data. Our analytic dataset includes background measures, including race (collected by the SCCS as White, Black, "other race," or Hispanic), gender, age at college entry, state residency status, dependency status for financial aid purposes, and high school graduation status. We are able to capture whether students work for pay while enrolled, their earnings during their first semester (as a proxy for the amount worked during initial enrollment), and their expected family contribution (EFC) for financial aid purposes. To understand how SAP failure impacts immediate persistence and achievement, we examine enrollment and GPA in the first term of the second year of college. Unfortunately, we are unable to capture students' degree-seeking status at college entry, and, for that reason, include a variety of outcomes. To explore the

implications of SAP for degree and transfer outcomes, we examine certificate and associate degree attainment and transfer to a four-year college within three years of initial enrollment.

State Contexts

The sampled state consistently falls within the lowest quintile in terms of average household-income across the nation (U.S. Census Bureau, 2014). The state also falls in the top quintile in terms of poverty rates, with 14 percent of the overall population and 19 percent of youth falling below the poverty line at the time of our final cohort (U.S. Census Bureau, 2007). The cost of college attendance is also among the lowest in the nation (U.S. Department of Education, 2013). The SCCS has substantially lower tuition rates than the national average. In 2007, the most recent cohort in our analytic sample, the maximum annual tuition and fees (16 credits per semester) at an SCCS college cost students approximately $1,475, about half of the national average of $2,708 for public two-year colleges (U.S. Department of Education, 2013).

The state's higher education governance structure includes a statewide governing board for the state's 17 public four-year institutions that is also responsible for statewide planning and data collection for all postsecondary institutions, including public community colleges and private four-year institutions. There is a separate governing board for the community colleges. The two boards collaborate on efforts that concern both sectors, such as aligning curriculum across the sectors, which enabled the creation of a statewide transfer articulation agreement.

An articulation agreement assures community college graduates admission to a state four-year college, though admission to a specific campus or program is not guaranteed. The articulation agreement stipulates that students who earn an associate degree and an overall GPA of at least 2.0, in addition to applying to a receiving institution by the stated deadline, will be

admitted to a four-year institution. Those denied admission to the specific institution are provided information on space availability and contacts at other public colleges in the system.

Descriptive Statistics

To describe the students in our sample and to demonstrate how they compare to first-time community college students across the nation, we present sample means for Pell and non-Pell community college students (based on Pell receipt at college entrance) in our pooled state administrative data and in the 2004/2009 Beginning Postsecondary Students Longitudinal Study (BPS:04/09) in Table 1. The BPS is a longitudinal study following a nationally representative sample of first-time college students, a subset of students from the National Postsecondary Student Aid Study of 2004.

[Insert Table 1 about here]

As Table 1 indicates, students in the SCCS sample are, on average, older than in a nationally representative sample. While both the BPS and the SCCS samples represent first-time college students, it is possible that there is variation in the age of first-time college students across states (perhaps reflecting differences in labor markets and college costs). In both the state and national data, half of Pell recipients are White, compared with approximately three quarters of non-Pell students. Approximately 70 percent of Pell recipients and 50 percent of non-Pell students are female. In our state sample, Pell recipients are more likely to have earned a high school diploma than non-Pell students.

Financially, the national and state samples differ across dependency status, employment status while enrolled, and EFC. Among Pell recipients, a similar proportion filed as financial dependents across the two samples—51 percent of students in the SCCS and 57 percent in the BPS. In the SCCS, the majority of the non-Pell students did not apply for financial aid and

therefore are "not dependent" for financial aid purposes (the dependency status is unknown and/or not applicable for two thirds of these students). Twenty-one percent fewer Pell recipients and 22 percent fewer non-Pell students worked for pay during the first semester in the SCCS than in the BPS. Finally, the EFC of non-Pell students in the national sample is much higher than in the SCCS. The financial differences are likely driven by the low household income in the sampled state. Furthermore, since the SCCS has substantially lower tuition than the national average, it may attract low-income students or result in fewer students applying for financial aid (almost 93 percent of non-Pell students did not file a FAFSA: see table notes in Table 1).

The non-negligible differences between our single-state sample and national data suggest that students in our sample are older and less likely to work for pay during college than a nationally representative sample. This may be due to the sluggish labor market in the sampled state. While our findings may not be directly generalizable to community college students across the nation, our focus on community college students in a state with a higher proportion of low-income citizens suggests that our results still offer important lessons regarding the impact of policies on students attending broad-access colleges, particularly in contexts where a higher than average percentage of the population lives in poverty.

Methods

We use several analytic approaches to examine the magnitude of SAP failure and its effects. First, we demonstrate the prevalence of failure to meet SCCS SAP standards using summary statistics. Next, we estimate the effects of SAP failure using two approaches. We compare the Pell recipients directly above and below the cutoff using a regression discontinuity design (RD). We also compare the effects of falling above and below the cutoff between those with and without federal aid in a difference-in-differences analysis (DD). The RD arguably

offers greater internal validity, evaluating the impact of SAP policy around the cutoff, where students are expected to either narrowly avoid or face penalty. The DD offers more external validity, capturing effects across a broader swath of the student population (not just those at the cutoff) and allowing us to net out any effects that may be due to more general SCCS academic performance standards that apply to all students, rather than SAP policy specifically.

Failure to meet SAP standards occurs when students fail to achieve a cumulative 2.0 GPA or obtain a cumulative .67 credit ratio of credits attempted/credits earned at the end of a given term, or to complete their program within 150 percent of the maximum stated time frame. Due to variation in maximum time frame across programs, we focus on the GPA and credit ratio requirements to study descriptive patterns. We then exploit the 2.0 GPA threshold, using cumulative GPA at the end of the first year of college (when students are evaluated for Pell renewal for the fall of their second year) to test the impacts of SAP on community college student persistence, transfer behavior, and degree attainment. The variation in cumulative GPA allows us to examine impacts around the cutoff, unlike the measure of credit ratios, which does not offer enough variation to estimate the effect of SAP failure (i.e., students do not attempt or accumulate many credits in the first year, producing a very lumpy distribution).

In our analyses, first-year cumulative GPA is an average of grades from fall, spring, and summer terms during year 1. For students who drop out, we use their GPA from enrolled terms to calculate cumulative GPA. Most students who fail to meet the GPA standard do so in the first semester—almost three quarters of those with a cumulative first-year GPA below 2.0 fall in this category—and about half appear to be discouraged from enrollment in a subsequent term. Almost all of the students who initially fell below 2.0 but reenrolled fail to meet SAP in the subsequent term (in fact, less than 1 percent of those who failed in the first term manage to

improve their standing). The remainder of students falling below the 2.0 mark in first-year cumulative GPA slip below the cutoff for the first time in their second term, and over half do not enroll in a subsequent semester. We estimate that, among the students who complete year 1 with a cumulative GPA below 2.0, less than 10 percent of students have received only one warning (this excludes those that dropped out immediately following their initial failure to meet SAP). Because some students are still in warning rather than facing ineligibility, our results may underestimate the effect of failing to meet SAP GPA standards on outcomes; some students we expect to lose Pell eligibility are actually still eligible and may not experience anticipated negative effects. Nonetheless, we choose to focus on students who fall below the GPA cutoff at the end of their first year, rather than relying on first-term GPA, to ensure that students have attempted enough credits to compute a reliable GPA. While first-term GPA could test the impact of receiving a warning, the measure does not offer enough variation to meet the assumptions necessary for our regression discontinuity design, described next.

Regression Discontinuity

To estimate the effect of SAP, we focus on students whose cumulative GPA falls near the cutoff—an approach referred to as local linear regression because the estimated RD impact is "local" to the cutoff, relying on a narrow bandwidth of GPAs (Hahn, Todd, & Van der Klaauw, 2001). The RD analysis relies on the assumption that, if student characteristics are evenly distributed across the "treatment" threshold (the 2.0 GPA cutoff), we can attribute any differences in outcomes across the threshold to the policy (Imbens & Lemieux, 2008). Using this approach, we estimate the difference between two regression functions at the cutoff—one estimated below and the other estimated above. The rationale for the local linear approach is that focusing on the subset of observations around the cutoff should allow us to more accurately

specify the functional form, reducing bias in estimation. To identify the optimal bandwidth around the cutoff, we use Calonico, Cattaneo, and Titiunik's (2014a, 2014b) robust bias-corrected confidence intervals and Imbens and Kalyanaraman's (2012) data-driven bandwidth selection. We use a bandwidth of 0.5 GPA points as our optimal bandwidth, based on Calonico et al.'s procedure, which results in a slightly more conservative estimate.

The basic model takes the form:

$$Y_i = \beta_0 + \beta_1(Below_i) + \beta_2(Distance_i * Below_i) + \beta_3(Distance_i * Above_i) + \varepsilon_i \quad (1)$$

where Y_i represents the outcome for student *i*, and β_1 provides an estimate of the effect of falling below the SAP cutoff on the outcome. *Below* is a binary indicator of whether or not the student's first-year GPA fell below the cutoff; *Above* indicates whether the student's GPA fell above the cutoff; *Distance* is the difference between the student's cumulative GPA and the 2.0 cutoff. ε_i is the residual component for student *i*. We test for sensitivity to bandwidth selection by running the basic model on the optimal bandwidth (model 1), a narrow bandwidth (half the optimal bandwidth; model 2), and a wide bandwidth (twice the optimal bandwidth; model 3).

To increase the precision of β_1 and to evaluate its sensitivity to additional covariates, we include institutional and cohort fixed effects, as well as individual-level measures, in additional models run on the optimal bandwidth. These models take the form:

$$Y_i = \beta_0 + \beta_1(Below_i) + \beta_2(Distance_i * Below_i) + \beta_3(Distance_i * Above_i) \quad (2)$$
$$+College\ FE + CohortFE + \beta_n X_i + \varepsilon_i$$

where *CollegeFE* represents a vector of institutional fixed effects (entered as a set of dummy variables indicating the institution initially attended, with one institution excluded), important because the financial aid officers responsible for implementing SAP policy are nested within institutions. Using institutional fixed effects also controls for other sources of between-college

variation, such as institution-specific academic interventions and initiatives that may influence student persistence and attainment. *CohortFE* is a vector of cohort fixed effects, a necessary inclusion because the implementation of campus policies (including timing of SAP notification) may vary over time, potentially producing varied effects. X_i represents a vector of individual-level covariates (and β_n the corresponding estimated effects on the outcome) including race, gender, age at initial enrollment, state residency status, dependency status for the purpose of financial aid, whether the student is working for pay, earnings during the first semester of college, and high school graduation status. Model 4 adds the institution and cohort fixed effects to model 1, ensuring that the RD is evaluated first within the institution and cohort before aggregating into β_1 to calculate the main effect. Model 5 adds the vector of covariates to model 4 to test for sensitivity to additional covariates. We interpret the final model (model 5) in the results section.

Testing the impact of SAP policy is not without its challenges. Given the nature of the state's grading system—in which only whole letter grades are awarded—and the fact that many community college students do not accumulate a lot of credits (making them more likely to only have one or two grades comprising their cumulative GPA), we find "heaping" in whole number GPAs across the distribution. The first panel in Figure 2 shows the GPA distribution among students in our sample. To estimate discontinuities using GPA as the running variable, we first must deal with "heaping-induced bias" in our state administrative data (Barreca, Lindo, & Waddell, 2011). Following the recommendations of Barreca, Lindo, and Waddell (2011; see also Barreca, Guldi, Lindo, & Waddell, 2011), we rely on "donut-RD" estimates, dropping observations at the whole number heaps that fall within our bandwidth. A drawback of this method is that it cannot estimate how SAP impacts students who tend to be observed in the

heaps. Still, the method is consistent with the usual motivation for RD and, based on simulations in Barreca, Lindo, and Waddell (2011), offers unbiased estimates.

[Insert Figure 2 about here]

While "lumpy" distributions for the running variable are often taken as evidence that there is manipulation of the running variable (McCrary, 2008), it is unlikely that students are able to manipulate their course grades to fall above the cutoff. However, we acknowledge that including the "whole" GPAs (i.e., 0.0, 1.0, 2.0, 3.0, and 4.0) introduces bias; students with whole GPAs may be less likely to persist because the presence of a whole GPA is often the result of fewer course credits contributing to the calculated average. The second panel of Figure 2 shows the distribution across the wide bandwidth after removing the heaps. The resulting smoothness of the donut GPS distribution across the cutoff and insignificant covariate discontinuities, described next, supports our decision. Ultimately, we find that using the donut-RD offers more conservative estimates.

To bolster support for an RD approach using students' first-year GPA as the running variable, we first test for discontinuities in covariates around the 2.0 cutoff. Regressions confirm that there are no systematic differences between Pell recipients who fell directly below and above the cutoff in terms of race, age, working for pay, holding a high school degree, and several other observable measures (results available in Table A1 of the Appendix). We find some evidence that women were more likely to fall below the cutoff than men, supporting our decision to control for gender in our covariate-adjusted model.

Difference-in-Differences

An institution's SAP policy must be at least as strict as the academic requirements for graduation, which, in the SCCS, is a 2.0 cumulative GPA. This means that our RD estimates

combine any general effects of failing institutional standards for academic performance with the specific effects of failing SAP standards for retaining financial aid. Therefore, we also apply a difference-in-differences (DD) approach to compare the effects of falling below the GPA cutoff among Pell Grant recipients (who face both SAP and general institutional consequences) and non-Pell students who do not receive any other form of federal or state aid (those who face only general institutional consequences). In essence, we use students who do not receive financial aid, which we refer to as "non-Pell students," as a control group for a comparison of the effects of falling below the SAP GPA standard. Among students who do not receive the Pell, only 3 percent received some other form of federal or state aid that would subject them to SAP standards (those students are not included in the analysis). Additional motivations for the DD approach are that it allows us to examine the effect of SAP for a wider range of students affected by the policy, rather than those right around the cutoff; moreover, the DD also provides greater power to detect effects than the RD.

The tradeoff is that the DD requires stronger assumptions about the relationship between first-year GPA and subsequent outcomes than the RD; namely, we must make a "parallel trends" assumption that outcomes vary by GPA in the same way for Pell and non-Pell recipients. In a community college sample, differences between Pell and non-Pell students are somewhat less concerning than in a four-year context, as research suggests that many community college entrants who may be eligible for federal grant aid do not file the FAFSA (Bettinger, Long, Oreopoulos, & Sanbonmatsu, 2012). More importantly, the parallel trends assumption does not require that Pell and non-Pell recipients have the same expected outcomes after controlling for GPA. This would be implausible, as we may expect (and indeed observe) that aid recipients have lower persistence than students not receiving aid, even when comparing students with the same

first-year GPA. Instead, the key assumption of the DD methodology is that whatever underlying differences exist between Pell and non-Pell students, these underlying differences are *fixed* or "parallel" across the range of GPAs. If we observe that Pell recipients above the GPA cutoff have moderately lower persistence than non-recipients with similar GPAs, but Pell recipients below the cutoff have *much* lower persistence than non-recipients with similar GPAs, the DD methodology will identify a negative effect of the policy.

We compare Pell and non-Pell students falling below and above the cutoff using the following basic model (model 1) on the optimal bandwidth:

$$Y_i = \beta_0 + \beta_1(Pell_i * Below_i) + \beta_2(Pell_i) + \beta_3(Below_i) + \varepsilon_i \quad (3)$$

where Y_i represents the outcome for student *i*, *Pell* is a binary indicator of whether the student received the Pell Grant at initial enrollment, and *Below* is a binary indicator of whether or not the student's GPA fell below the cutoff at the end of year 1. Coefficient β_1, our estimate of interest, represents the effect of falling below the 2.0 GPA threshold for students who entered college with a Pell versus those who did not (the difference-in-differences between students who fall below and above the 2.0 GPA threshold at each Pell condition). β_2 is the main effect of Pell on the outcome; β_3 is the main effect of falling below the cutoff on the outcome. ε_i is the residual component for student *i*.

We conduct a series of regression analyses, adding institutional and cohort fixed effects and individual-level covariates to the basic model:

$$Y_i = \beta_0 + \beta_1(Pell_i * Below_i) + \beta_2(Pell_i) + \beta_3(Below_i) + College\ FE \quad (4)$$
$$+CohortFE + \beta_n X_i + \varepsilon_i$$

where *CollegeFE* represents a vector of institutional fixed effects and *CohortFE* represents a vector of cohort fixed effects. Coefficient β_1 represents the effect of falling below the 2.0 GPA

threshold for students who entered college with a Pell versus those who did not in a given institution and cohort. X_i represents a vector of individual-level covariates. ε_i is the residual component for student i.

Model 2 includes institutional and cohort fixed effects, but does not yet include student-level covariates. Model 3 adds the distance of students' GPA from the cutoff ($Distance_i$) to model 2. In order to capture the variation in slopes across Pell status observed above and below the GPA threshold, model 4 adds the student's GPA relative to the cutoff, $Distance_i * Below_i$, to the previous model. Our most comprehensive model, model 5, adds control measures. We interpret model 5 in the results section.

Results

In this section, we describe the results of our analyses, beginning with the descriptive patterns of interest in research question (RQ) 1. Then, we turn to the results from the regression discontinuity (RQ2) and the DD analysis (RQ3).

The Prevalence of SAP Failure and its Descriptive Relationship to Persistence (RQ 1)

Our administrative data allow us to examine descriptive patterns in SAP failure across both the SAP credit requirement and the GPA requirement. We are able to look at these measures by academic term, as they are evaluated in the SCCS, rather than by academic year. We begin by plotting the rates of failure to meet SAP requirements among all SCCS students across entry cohorts in Figure 3, which shows students falling below the GPA and credit ratio standards, along with a combined estimate of falling below SAP overall in the first year.

[Insert Figure 3 about here]

The figure indicates increasing rates of failure over time for the GPA standard, consistent with the national data for community colleges. For the years in which we have comparable data,

the levels of failure for the GPA requirement are slightly higher in the SCCS sample than we observed in the national data in Figure 1. This may simply reflect regional variation, or it may be due to the fact that our administrative data represent true first-time entrants in their first term, while the NPSAS data include all students with first-year academic standing. Figure 3 also indicates that at least as many students are failing the credit standard as are failing the GPA standard. Of course, there is some overlap between the two requirements, but accounting for students who fail either of the two standards raises the overall first-term SAP failure rate substantially, from 25 percent to 42 percent for students who entered in the fall of 2008.

Figure 4 plots the prevalence of SAP failure and enrollment patterns among Pell recipients over the first 3 years of college. To the extent that SAP failure has a consequence, it likely shows up as dropout rather than as persistence without financial aid. The figure demonstrates that SAP failure is most prevalent in the first term of enrollment, where approximately 40 percent of Pell recipients failed SAP in the first term. By the fall of the second year, the rate is more than halved (to 16 percent), and by the fall of the third year, the rate is halved again (to a little over 8 percent of Pell entrants). The prevalence of SAP failure declines over time not because students improve their GPAs overall (though, of course, some do) but because many students simply do not reenroll, and those who fail SAP are disproportionately likely to drop out (29 percent of Pell recipients who fail SAP in the first term do not return in the spring, compared with just 13 percent of Pell recipients who meet SAP standards). The students who return in subsequent semesters are thus positively selected and less at risk of SAP failure to begin with.

[Insert Figure 4 about here]

We find that, among SCCS students who remained enrolled after failing SAP, the majority continued to receive Pell Grants, likely a result of successful appeal. In the fall term of students' second year, 73 percent of students who failed to meet SAP standards and remained enrolled still receive the Pell Grant (though, most Pell recipients who failed SAP dropped out by this point—approximately 60 percent). Per institutional policy, students have one semester to improve their SAP status while on probation-by-appeal. The rate of Pell retention falls significantly after that point. By fall of the third year, 3 percent of Pell entrants who continually fall below SAP standards appear to retain the Pell.

While students who fail to meet SAP requirements are more likely to drop out, the difference in retention rates between those who do and do not fail SAP has no obvious causal interpretation. Figure 5 provides preliminary evidence for both our RD and DD methods, plotting GPA and persistence over time for Pell and non-Pell students. Panel A looks at fall-to-spring persistence by first-term GPA, while panel B looks at fall-to-fall persistence by cumulative first-year GPA (those who drop out during the first year remain in the sample—their GPA at exit is carried forward).

[Insert Figure 5 about here]

Interestingly, panel A indicates that Pell recipients persist to the spring term at higher rates than non-Pell students across the GPA spectrum. We do not see any strong changes in patterns of persistence around the 2.0 GPA cutoff. Panel B looks quite different. Here, students below the 2.0 GPA threshold who entered with Pell Grants appear to have lower rates of persistence to the subsequent fall term than non-Pell entrants with similar GPAs; above the threshold, the two groups persist at similar rates. It is plausible that the SAP rules embedded in the Pell Grant program explain these patterns: The threat of losing the Pell due to SAP rules may

discourage students who entered with the Pell from continuing to the fall of their second year. We dig deeper into these patterns in our subsequent analyses.

Effects of Falling Below 2.0 around the Cutoff: Regression Discontinuity Results (RQ 2)

The RD analysis is our first effort to more closely examine the potential discouragement effect of the SAP standards. Table 2 shows the effect of failing to meet the GPA standard at the end of the first year of college, when students in the warning period should face the threat of Pell loss, on reenrollment in the second year of college, first-term GPA from the second-year, certificate and associate degree attainment, and transfer to a four-year college. While we might expect to see an immediate negative impact on persistence (whether from Pell loss or discouragement due to the threat of Pell loss), the RD results do not support that hypothesis. We find null effects across all four outcomes. Our results are imprecise, and small negative effects cannot be ruled out. To provide additional insight into the distribution of student outcomes across GPA, Figures A1-A5 in the Appendix present the RD results graphically.

[Insert Table 2 about here]

Effects of Falling Below 2.0 for Pell vs. Non-Pell Students: Difference-in-Differences Results (RQ 3)

The DD results, presented in Table 3, expand on the patterns observed in Figure 5, panel B. In contrast with the RD results (which focus on Pell recipients only), the DD suggests a negative effect of SAP on persistence for Pell recipients. Pell students who fail to meet the GPA standard for SAP are nearly 5 percentage points less likely to persist to the second year of college than non-Pell students who fail to earn a 2.0 GPA but do not face SAP standards ($p<.001$).

[Insert Table 3 about here]

While SAP policy appears to negatively impact early college persistence, we find no impact on first-year GPA and small positive effects on associate degree attainment and transfer to a four-year college within three years of college entrance. Pell recipients who failed to meet SAP standards in their first year were approximately 1 percentage point more likely to earn an associate degree ($p<.05$) and 3 percentage points more likely to transfer to a four-year college than their non-Pell peers who also fell below the 2.0 mark in their first year ($p<.01$). Figures A6-A10 in the Appendix provide graphics of the DD results.

Taking a closer look at the distribution in the graphic representation, the gap between Pell recipients and non-aid recipients does indeed appear larger above and smaller below the threshold (see Figure A9 for the associate degree results and, to a lesser extent, Figure A10 for transfer in the Appendix). However, the probability that Pell recipients will earn an associate degree in three years is so low that students falling below a 2.0 "bottom out" (i.e., they have zero probability of earning a degree and cannot go lower), giving the illusion, when merely looking at the difference in effects, that Pell students are narrowing in on their non-Pell counterparts. This implies that the positive effects for long-term outcomes may be due to floor effects, rather than incentivizing effects. Similar patterns emerge for the transfer outcome, but students appear to bottom out at a probability of .15 rather than 0. We suspect that this may be a result of students being concurrently enrolled at a community college and four-year college, as it seems unlikely that a non-negligible portion of students with cumulative GPAs below 2.0 are transferring to four-year colleges. Unfortunately, we cannot confirm concurrent enrollment with our data.

Discussion

In this paper, we provide evidence regarding the prevalence and consequences of students' failure to meet SAP standards using state administrative data from community colleges.

The national trends we highlight in the introduction (refer to Figure 1) suggest that failure to meet academic standards is not a Pell recipient–specific phenomenon. Yet while many students fail to meet minimum academic standards, SAP policy targets undergraduates from America's most disadvantaged families (most Pell recipients come from families earning under $40,000 annually) who are in greatest need of support in order to attain a degree and who stand to benefit the most from degree attainment (Baum & Payea, 2013). Students who rely on federal financial aid to attend college must meet academic standards more quickly than students who can pay for college out of pocket, particularly because academic probation in the SCCS has no consequence for enrollment, only for graduation. If students become ineligible for the Pell Grant, their net price of attending college increases by at least the size of their Pell Grant (or more, if they received other financial aid tied to Pell eligibility).

Based on our descriptive results from SCCS, many Pell recipients—approximately 40 percent—risk losing aid due to SAP failure. While several forms of information about SAP guidelines were provided to SCCS students, it is impossible to confirm from our data that students were aware that additional academic requirements are imposed on financial aid recipients. If students do not become aware of the requirements until after notification of failure at the end of the first semester, prior grades still contribute to the cumulative GPA calculation at the end of the first year. Notification of SAP in the middle of year 1 could be too late for some students to turn it all around.

Our descriptive statistics and DD results offer some evidence in support of a discouragement effect. Pell recipients experienced greater negative effects of falling below 2.0 than their non-Pell peers who were subject to only institutional general academic standards. However, the RD estimates—which are null—do not bolster support for the discouragement

theory, offering little evidence of discontinuous outcome differences for students just above and below the GPA cutoff. As we move further below the cutoff, as in the DD analysis, the discouragement effect from failing to meet the standard may grow, producing the negative impacts on persistence observed in Table 3. Getting back over the 2.0 threshold becomes more difficult the farther students fall below it. The impacts of falling below are larger for Pell students than non-Pell students because of the threat of financial aid loss.

At the same time, we cannot entirely rule out the theory that SAP policy may incentivize effort among some students. We are hesitant to overemphasize the significant positive impacts of SAP on degree attainment and transfer from the DD, particularly given the minimal impact on second-year GPA and potential floor effects described above. The direction of the GPA estimate leaves the debate over incentivizing effects open, as we do not have evidence to the contrary. At the very least, SAP policies do not appear to have a strong, average incentive effect.

Adjudicating between the null effects in the RD analyses and the statistically significant effects in the DD is complicated, as each set of analyses comes with its own assumptions. The RD provides more causally rigorous estimates, but also has two limitations. First, it estimates the joint effect of failing to meet SAP and general academic standards for students surrounding the cutoff. Given the nature of federal policy, SAP policies often coincide with an institution's academic standards. This may mean that the RD provides more reliable estimates because the two policies occur jointly in practice (though, it is possible that general academic standards have sharper teeth in other institutional settings than they do in SCCS), but the RD is not necessarily isolating the impact of SAP failure. Second, the RD focuses on a narrower sample of students—capturing the impact local to the cutoff, even though our theoretical framework strongly suggests that responses to the policy may vary depending upon how far students are from the cutoff.

The DD estimates, on the other hand, attempt to isolate the impact of SAP policy, which is of interest in examining whether threat of Pell loss, rather than of general academic failure, impacts student success. The DD also has the advantage of estimating effects for a wider range of students. The DD analysis, however, requires additional assumptions that come with using non-Pell students as a control group. Specifically, the DD design requires that any between-group differences must be fixed/parallel across the range of GPAs considered. If Pell recipients' outcomes are *more* different from non-recipients' outcomes below the cutoff than above the cutoff, for reasons that have nothing to do with SAP policy, the DD estimate will be biased. Still, given trends of imperfect take-up of Pell grant among the Pell-eligible population, we think it is plausible that many Pell-eligible students exist within the non-aided student comparison group. This would suggest that the two groups may not be so dissimilar after all, but, ultimately, the plausibility of the parallel trends assumption is impossible to fully verify.

An additional consideration in interpreting our results concerns our focus on the GPA component. Focusing on the GPA component of SAP, while methodologically necessary, means that some students included as "passing SAP" may actually fail to meet the credit ratio standard. We estimate that, of the students in our sample who appear to meet the GPA standard, about 11% of them are failing to meet the credit ratio standard, which would mean a portion of our control group is also subject to financial aid ineligibility. This provides another reason why our analysis may underestimate the effects of failing to meet SAP.

Data limitations preclude some additional analyses that would be of interest. For example, while the SCCS administrative data provide us with measures of students' grades, credits, and outcomes, we do not have access to their SAP status, as determined by their financial aid office. It appears that many students facing Pell loss due to SAP failure initially avoid

negative consequences by appealing. While students cannot delay negative effects for too long (an appeal only provides one additional semester to meet SAP standards, and all of the standards are cumulative), it may dampen the negative effects we observe. Additional data on the status awarded by the financial aid office could better illuminate patterns of effects. Furthermore, measures of prior academic achievement would be useful in examining effect heterogeneity, but are not available in the SCCS data.

Conclusion

This paper offers the first attempt to uncover the effects of SAP standards. While the analytic approaches we use are sound for the study's goals, more research is necessary to further investigate the effects of SAP. Ideally, a randomized controlled trial, assigning some federal aid recipients to meet SAP standards, while waiving the standards for others, would be an ideal model for testing the effects of tying academic standards to need-based financial aid. Such a study seems unlikely without significant collaboration between researchers, institutions, and government officials, since SAP standards are federally mandated.

While it is possible that SAP improves performance of the students who persist despite initial SAP failure, we do not find clear evidence that it incentivizes effort using either identification strategy. If anything, our results suggest that weeding out, where students are discouraged by information that they failed to meet academic standards by dropping out, is a more likely response. It is concerning that SAP policy may push Pell recipients—already notably at greater risk of drop out than more affluent students—out of college before they have a chance to adjust to the academic demands of college life.

Even if SAP standards do not incentivize students to improve their grades, it is unlikely that policymakers would eliminate performance-based standards—both due to the need to keep

the Pell efficient and the potential political backlash of making need-based federal aid seem like a handout. Performance-based standards may make policies more politically palatable. According to the America's College Promise fact sheet, President Obama's "free community college" proposal sets performance standards even higher than SAP for federal aid (Whitehouse.gov, 2015). To remain eligible, community college students would need to maintain a 2.5 GPA, in addition to remaining enrolled half time and demonstrating "steady progress" toward a degree (Whitehouse.gov, 2015). In 2012, 37% of first-year CC students earned below a 2.5, suggesting that a substantial portion of initial America's College Promise enrollees would quickly lose eligibility (authors' calculations, NPSAS 2012). Although our research cannot speak to the ideal placement for standards, it highlights the urgent need to address the fact that many students immediately fail to meet standards. While academic preparation is a piece of the puzzle, it's likely additional student support services could help students improve their outcomes and would be necessary for a program like America's College Promise to significantly increase the number of students earning a college credential.

Drawing strong conclusions from our single-state estimates would be premature, especially given that the RD does not corroborate our statistically significant DD results, which have more questionable internal validity. This study serves as a point of departure from which policy scholars should continue exploring the implications of tying need-based aid to performance standards. The sheer magnitude of SAP failure—which is not merely concentrated in community colleges, but appears just as common in public four-year institutions—indicates that researchers interested in higher education policy should no longer ignore the role that SAP may play in moderating the effects of federal aid.

References

Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics*, 11, 48–83.

Angrist, J., Autor, D., Huson, S., & Pallais, A. (2014). *Leveling up: Results from a randomized evaluation of post-secondary aid*. (NBER Working Paper No. 20800). Cambridge, MA: National Bureau of Economic Research.

Attewell, P., Heil, S., & Reisel, L. (2011). Competing explanations of undergraduate noncompletion. *American Educational Research Journal, 48*(3), 536–559.

Bettinger, E., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R block FAFSA experiment. *Quarterly Journal of Economics 127*(3), 1205-1242.

Barreca, A. I., Guldi, M., Lindo, J. M., & Waddell, G. R. (2011). Saving babies? Revisiting the effect of very low birth weight classification. *Quarterly Journal of Economics, 126*(4), 2117–2123.

Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2011). *Heaping-induced bias in regression-discontinuity designs*. (NBER Working Paper No. 17408). Cambridge, MA: National Bureau of Economic Research.

Barrow, L., Richburg-Hayes, L., Rouse, C. E., & Brock, T. (2014). Paying for performance: The education impacts of a community college scholarship program for low-income adults. *Journal of Labor Economics, 32*(3), 563-599.

Baum, S., & Payea, K. (with Kurose, C.). (2013). *Trends in student aid 2013*. Retrieved from the College Board website: http://trends.collegeboard.org/sites/default/files/student-aid-2013-full-report-140108.pdf

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117(3), 871–915.

Bennett, W., & Grothe, B. (1982). Implementation of an academic progress policy at a public urban university: A review after four years. *Journal of Student Financial Aid, 12*(1), 33–39.

Bettinger, E. (2004). How financial aid affects persistence. In C. M. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 207–238). Chicago, IL: University of Chicago Press.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014a). Robust data-driven inference in the regression-discontinuity design. *Stata Journal, 14*(4), 909-946.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014b). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica, 82*(6), 2295-2326.

Carruthers, C. K., & Ozek, U. (2013, February). *Losing HOPE: Financial aid and the line between college and work*. National Center for Analysis of Longitudinal Data in Education Research Working Paper 91.

Casey, M., Cline, J., Ost, B., & Qureshi, J. (2015, February). *Academic probation, student performance and strategic behavior*. Paper presented at the annual meeting for the Association for Education Finance and Policy, Washington, DC.

Castleman, B., & Long, B. T. (2013). *Looking beyond enrollment: The causal effect of need-based grants on college access, persistence, and graduation* (NBER Working Paper No. 19306). Cambridge, MA: National Bureau of Economic Research.

Cohodes, S. R., & Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' Adams Scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics, 6*(4), 251-85.

Cornwell, C., Mustard, D. B., & Sridhar, D. J. (2006). The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Program. *Journal of Labor Economics*, 24(4), 761 – 786.

Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student Responses to Merit Scholarship Retention Rules. *Journal of Human Resources*, 40, 895–917.

Denning, J. (2014, December). *College on the cheap: Costs and benefits of community college*. University of Texas: Working paper.

Dynarski, S. M. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. *American Economic Review, 93*(1), 279–288.

Dynarski, S. M. (2008). Building the Stock of College-Educated Labor. *Journal of Human Resources*, 43(3): 577-610.

Goldrick-Rab, S. (2013). *Increasing the impact of student financial aid: Three recommendations for financial aid administrators*. Washington, DC: HCM Strategists.

Goldrick-Rab, S., Kelchen, R., Harris, D., & Benson, J. (Forthcoming). Reducing income inequality in higher education: Experimental evidence on the impact of financial aid on college completion. *American Journal of Sociology*.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201–209.

Henry, G.T., Rubenstein, R., & Bugler, D.T. (2004). Is HOPE enough? Impacts of receiving and losing merit-based financial aid. *Educational Policy*, 18(5), 686-709.

Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies, 79*(3), 933–959.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615–635.

Kane, T. J. (1995, July). *Rising public college tuition and college entry: How well do public subsidies promote access to college?* (NBER Working Paper No. 5164). Cambridge, MA: National Bureau of Economic Research.

Lindo, J. M., Sanders, N. J., & Oreopoulos, P. (2010). Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics, 2*(2), 95–117.

Manski, C. F. (1989). Schooling as experimentation: A reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review*, *8*(4), 305–312.

Martorell, P., McCall, B. P., & McFarlin, I. (2014, September). *Do public tuition subsidies promote college enrollment? Evidence from community college taxing districts in Texas*. US Census Bureau Center for Economic Studies Paper No. CES-WP- 14-32.

Mayer, A. K., Patel, R., Rudd, T., & Ratledge, A. (with Blake, S.). (2015). *Designing scholarships to improve college success: Final report on the Performance-based scholarship demonstration.* New York: MDRC.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*(2), 698-714.

McNair, E., & Taylor, S. E. (1988). Satisfactory academic progress standards: Jeopardizing efforts toward educational equity? *Journal of Student Financial Aid, 18*(1), 10–17.

Richburg-Hayes, L., Brock, T., LeBlanc, A., Paxson, C., Rouse, C. E., & Barrow, L. (2009). *Rewarding persistence: Effects of a performance-based scholarship program for low-income parents*. New York, NY: MDRC.

Satisfactory Academic Progress, 34 C.F.R. § 668.34 (2013). Electronic Code of Federal Regulations. Retrieved from http://www.ecfr.gov/cgi-bin/text-idx?c=ecfr&tpl=/ecfrbrowse/Title34/34cfr668_main_02.tpl

Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources, 46*(3), 614–646.

Student Eligibility, 20 u.s.c. 1091 H.E.A. § 484 (2013). Retrieved from http://www.house.gov/legcoun/Comps/HEA65_CMD.pdf

U.S. Census Bureau. (2014). *Median Household Income (In 2013 Inflation-Adjusted Dollars) by State Ranked from Highest to Lowest Using 3-Year Averages*. Retrieved from http://www.census.gov/hhes/www/income/data/statemedian/

U.S. Census Bureau. (2007). *Small Area Income and Poverty Estimates: State and County Estimates for 2007*. Retrieved from https://www.census.gov/did/www/saipe/data/statecounty/data/2007.html

U.S. Department of Education, National Center for Education Statistics. (2013). *Digest of education statistics, 2012* (NCES 2014-015), Chapter 3, Table 381. Retrieved from http://nces.ed.gov/programs/digest/d12/tables/dt12_381.asp

Welbeck, R., Ware, M., Cerna, O., & Valenzuela, I. (with Ratledge, A., & Boynton, M.). (2014). *Paying it forward: A technical assistance guide for developing and implementing performance-based scholarships*. New York, NY: MDRC.

Whitehouse.gov. (2015). *Fact sheet - White House unveils America's College Promise proposal: Tuition-free community college for responsible students*. Retrieved from: https://www.whitehouse.gov/the-press-office/2015/01/09/fact-sheet-white-house-unveils-america-s-college-promise-proposal-tuitio

Tables

Table 1

Descriptive Statistics: National and State Samples of Community College Students

	BPS		SCCS	
Variable	**Non-Pell**	**Pell**	**Non-Pell**	**Pell[a]**
White	0.735	0.517	0.738	0.513
	(0.442)	(0.499)	(0.440)	(0.500)
Black	0.090	0.271	0.150	0.379
	(0.286)	(0.445)	(0.357)	(0.485)
Asian	0.044	0.037		
	(0.206)	(0.188)		
Other race	0.105	0.149	0.072	0.082
	(0.306)	(0.356)	(0.259)	(0.274)
Hispanic	0.135	0.164	0.040	0.026
	(0.341)	(0.370)	(0.195)	(0.016)
Female	0.526	0.705	0.508	0.674
	(0.499)	(0.456)	(0.500)	(0.469)
Age	21.875	22.257	23.881	23.544
	(7.849)	(6.884)	(9.156)	(7.977)
State resident	0.954	0.972	0.876	0.951
	(0.210)	(0.165)	(.330)	(0.216)
Dependent	0.668	0.567	0.101	.514
	(0.471)	(0.495)	(.301)	(0.500)
Work for pay	0.786	0.694	0.570	0.480
	(0.410)	(0.461)	(.495)	(0.500)
First semester earnings ($)			181.935	133.89
			(275.194)	(180.32)
High school graduate	0.879	0.847	0.797	0.887
	(0.326)	(0.359)	(.402)	(0.317)
Expected family contribution ($)	11632.41	899.42	3295.15[b]	758.64
	(14,314.19)	(1,163.75)	(7,236.70)	(1231.24)
N	2,900	1,620	104,545	42,835

Notes: Table shows averages for Pell and non-Pell community college entrants from the BPS 04/09 (N= 4,520) and state administrative data from fall entry cohorts from 2002-2007 (N= 147,380). We used survey weights for the BPS data, which represents 1,144,570 students nationally, including a subpopulation of 299,810 Pell recipients and 844,760 non-Pell students. In correspondence with National Center for Education Statistics requirements, all sample sizes for the BPS are rounded to the nearest 10.

[a] The federal Pell Grant is, by far, the most prevalent form of financial aid in the sample, held by almost 30 percent of SCCS entrants.

[b] While the EFC is available for all non-Pell students in the NPSAS, it is only available for students who filed the FAFSA in the SCCS data. Only 7,676 of the 104,545 non-Pell students in the sample have data on their EFC, suggesting that only 7% of non-Pell SCCS students filed the FAFSA.

Table 2

SAP Impact on College Outcomes: Regression Discontinuity Effects within 3 Years of Entry

Outcomes	**M1**	**M2**	**M3**	**M4**	**M5**
Persist into year 2	0.026	0.071	0.03	0.038	0.028
	(0.024)	(0.039)	(0.026)	(0.023)	(0.024)
First-term, second- year GPA	0.011	0.059	0.004	0.011	0.011
	(0.038)	(0.064)	(0.041)	(0.038)	(0.038)
Certificate	0.003	0.002	0	0.003	0.004
	(0.005)	(0.009)	(0.006)	(0.005)	(0.005)
Associate degree	-0.009	-0.01	-0.015	-0.008	-0.008
	(0.007)	(0.013)	(0.008)	(0.007)	(0.007)
Transfer to four-year	-0.018	0.002	-0.024	-0.014	-0.016
	(0.020)	(0.032)	(0.021)	(0.020)	(0.019)
N	9,064	4,043	18,994	9,064	9,064
Specifications					
Bandwidth	.5	.25	1.0	.5	.5
Form	Linear	Linear	Quadratic	Linear	Linear
Fixed effects	No	No	No	Yes	Yes
Covariates	No	No	No	No	Yes

Notes: Table presents RD estimator coefficients, with standard errors in parentheses. The functional form for the optimal (.5) and narrow (.25) bandwidth is local linear, while the functional form for the wide (1.0) bandwidth is local quadratic. "Fixed effects" refer to institutional and cohort fixed effects, included in the models as a series of dummy variables with the reference eliminated.

* p<.05; **p<.01; ***p<.001

Source: SCCS administrative data.

Table 3

SAP Impact on College Outcomes: Difference-in-Differences Results

Outcomes	**M1**	**M2**	**M3**	**M4**	**M5**
Persist into year 2	-0.050***	-0.048***	-0.048***	-0.047***	-0.046***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
First-term, second-year GPA	0.010	0.013	0.011	0.012	0.015
	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
Certificate	0.000	0.001	0.001	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Associate degree	0.009*	0.010*	0.010*	0.010*	0.010*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Transfer to four-year	0.030**	0.033**	0.033**	0.033**	0.030**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Specifications:					
Fixed effects	No	Yes	Yes	Yes	Yes
Covariates					
Distance	No	No	Yes	Yes	Yes
*Distance*Below*	No	No	No	Yes	Yes
Controls	No	No	No	No	Yes

Notes: N= 31,768. Table presents DD estimator coefficients, with standard errors in parentheses. Each model is performed on the optimal (.5) bandwidth with a linear functional form. "Fixed effects" refer to institutional and cohort fixed effects, included in the models as a series of dummy variables with the reference eliminated. To put results into context, the average for each outcome among the control group (non-Pell students who fail to meet SAP): persistence=0.521, second-year GPA= 1.660, certificate=0.011, associate=0.021, transfer=0.192.

* p<.05; **p<.01; ***p<.001

Source: SCCS administrative data.

Figures

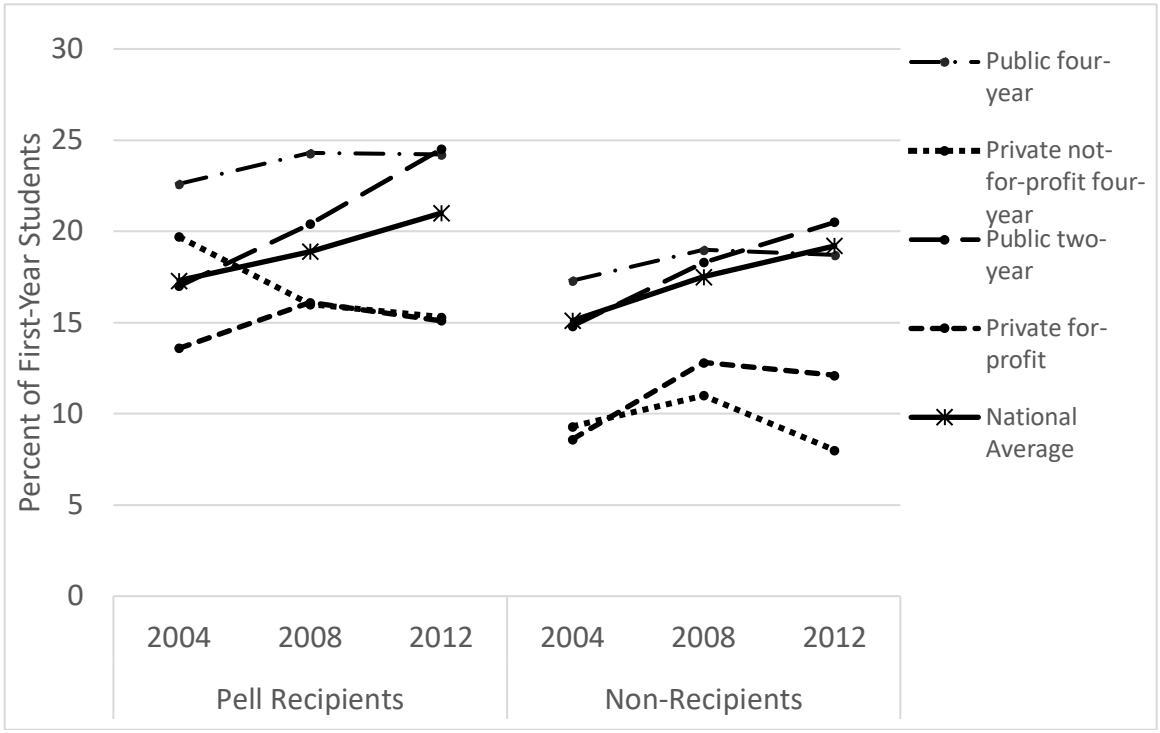

Figure 1. National Trends in SAP GPA Failure During the First Year of College: Averages across Institution Type and Pell Status

Notes: Figure displays percentage of students failing to achieve a 2.0 or higher GPA in the given academic year, estimated using National Postsecondary Student Aid Study (NPSAS) 2004, 2008, 2012 data on first-year-equivalent students. Federal SAP regulations require institutions to evaluate SAP for all federal aid recipients at the end of each academic year, where they must meet a 2.0 by the end of the second year. The left side of the x-axis displays average rates of failure for Pell recipients and the rightmost side displays averages for students who do not receive the Pell.

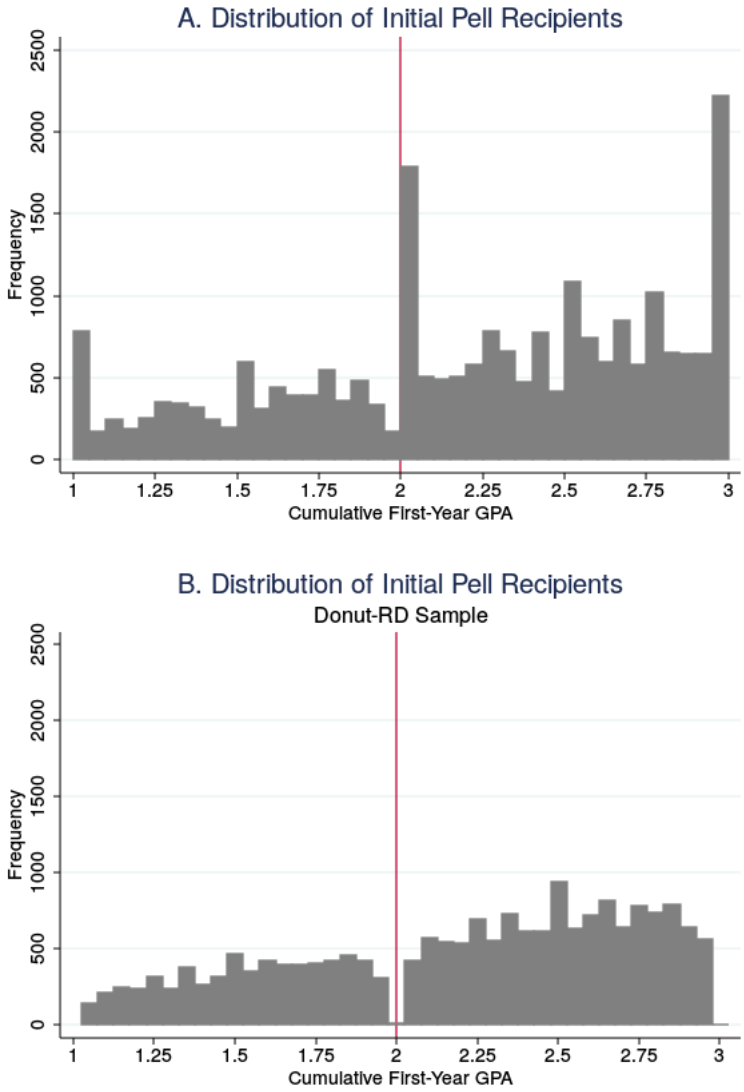

Figure 2. Distribution of Pell Grant Recipients: Changes in Distribution when Eliminating Heaping

Notes: Panel A shows the distribution of student GPAs across the wide bandwidth, within 1 grade point of the cutoff (N=20,567, wide bandwidth). There is heaping in the bin containing the 2.0 cutoff (along with the other .0-containing bins). Panel B shows the distribution of student GPAs used for donut-RD analysis, where students with "whole" GPAs are removed from analytic sample (N=18,994, wide bandwidth).

Source: SCCS administrative data.

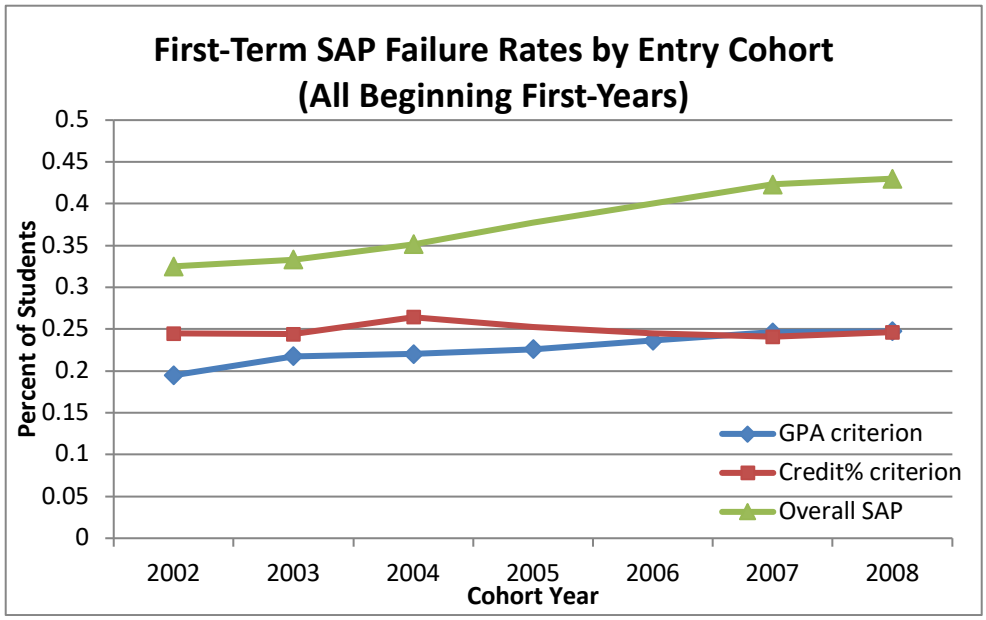

Figure 3. First-Term SAP Failure Rates by Entry Cohort (All Beginning First-Years)
Notes: N= 147,380. The figure presents the percent of first-time SCCS students who fail to meet SAP-G, SAP-C, and, subsequently, SAP-overall standards within their first term of enrollment. Due to problems with the "credits attempted" measure in 2005 and 2006, we were unable to calculate students' credit ratio, and therefore the overall SAP estimate, for 2005-2006.
Source: SCCS administrative data.

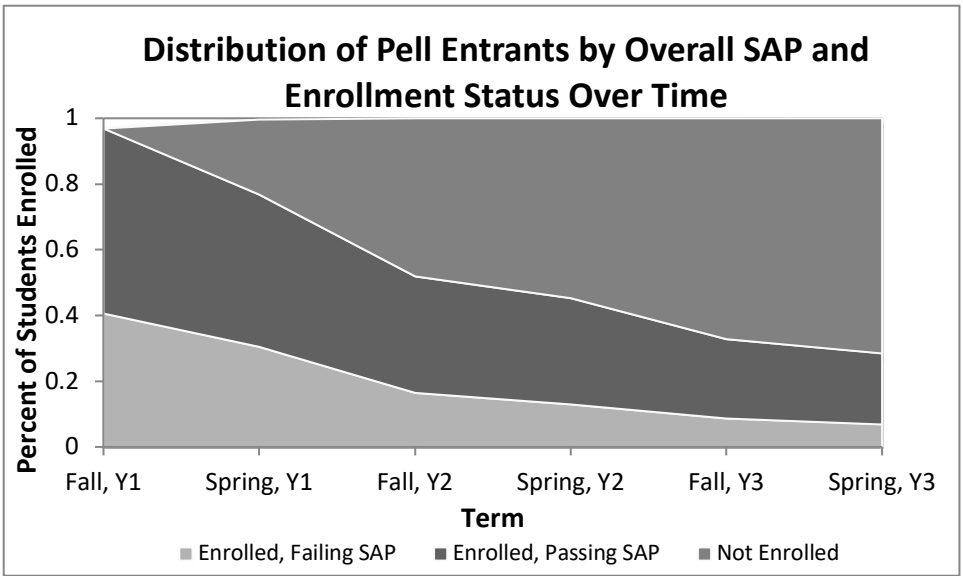

Figure 4. Enrollment by SAP Status across Academic Terms

Notes: N=42,835. The figure presents Pell entrants' enrollment and SAP status over time (excluding summer terms, when fewer students are enrolled). In the first term, a small percentage of enrolled students have no valid GPA. Otherwise, these categories are mutually exclusive and should add to 100 percent.

Source: SCCS administrative data.

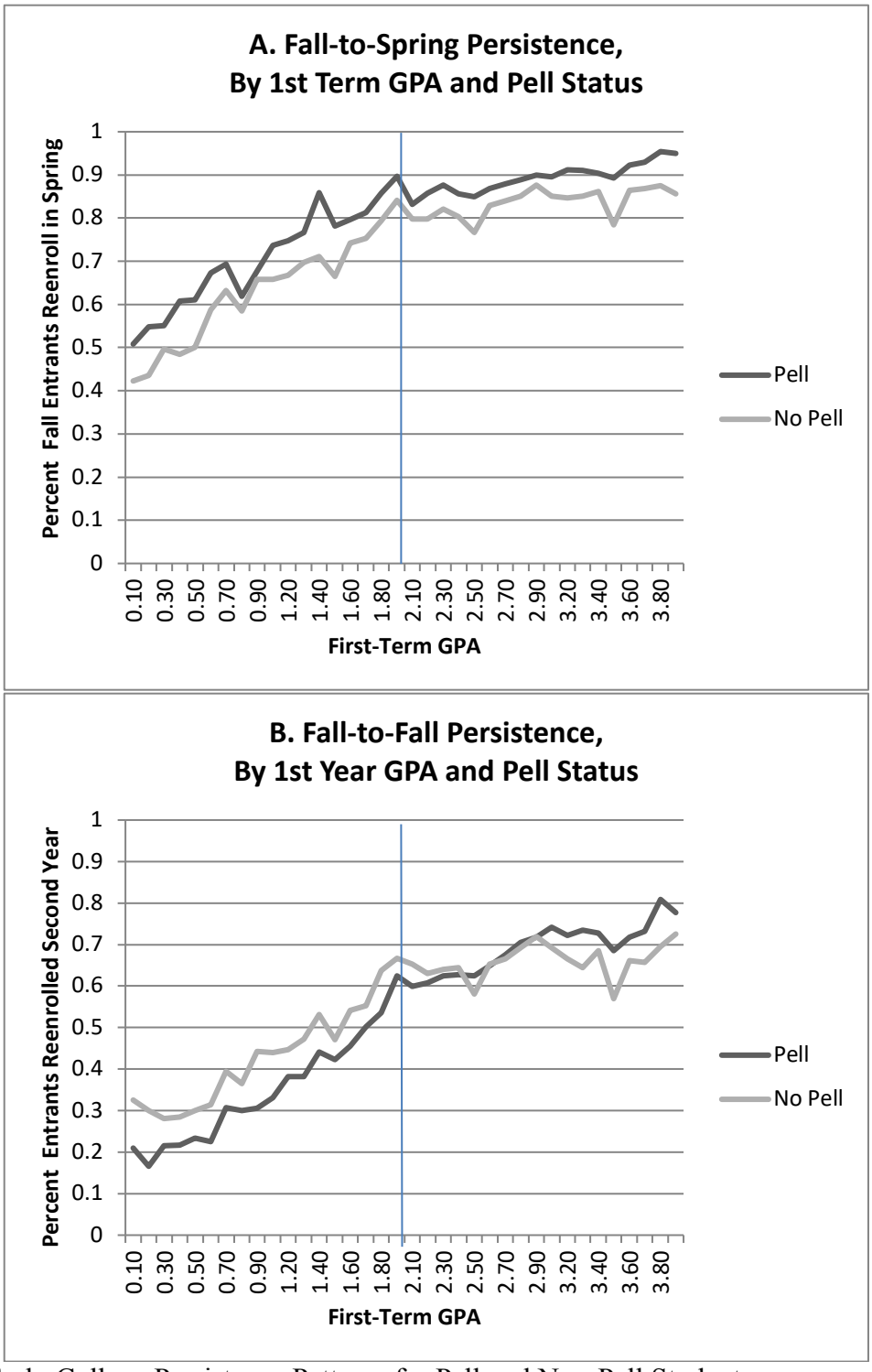

Figure 5. Early College Persistence Patterns for Pell and Non-Pell Students
Notes: N= 147,380. The percent of enrolled is presented within .05 GPA bins. In both panels, data for whole GPAs (1.00, 2.00, 3.00, 4.00) are hidden.
Source: SCCS administrative data.